AD-A058 884 DELAWARE UNIV NEWARK DEPT OF STATISTICS AND COMPUTER--ETC F/G 12/1
EFFICIENT ALGORITHMIC SOLUTIONS TO EXPONENTIAL TANDEM QUEUES WI--ETC(U)
JUN 78 G LATOUCHE, M F NEUTS AFOSR-77-3236
UNCLASSIFIED TR-78/12 AFOSR-TR-78-1357 NL

1 OF 1
ADA
058884

END
DATE
FILMED
11-78
DDC

INSTITUTE FOR
MATHEMATICAL SCIENCES

University of Delaware
Newark, Delaware

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78 - 1357	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EFFICIENT ALGORITHMIC SOLUTIONS TO EXPONENTIAL TANDEM QUEUES WITH BLOCKING		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Guy Latouche and Marcel F. Neuts		8. CONTRACT OR GRANT NUMBER(s) AFOSR 77-3236
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Delaware, Department of Statistics & Computer Science, Newark, Delaware 19711		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE June 1978
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Queueing systems, blocking, buffer models, computational probability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Stable queueing systems consisting of two groups of servers, having exponential service times, placed in tandem and separated by a finite buffer, are shown to have a steady-state probability vector of matrix-geometric form. The queue is stable as long as the Poisson arrival rate does not exceed a critical value, which depends in a complicated manner on the service rates, the numbers of servers in each group, the size of the intermediate buffer and the unblocking rule followed when system becomes blocked. The

next page

DD FORM 1 JAN 73 1473

UNCLASSIFIED

20. Abstract continued.

critical input rate is determined in a unified manner.

For stable queues, it is shown how the stationary probability vector and other important features of the queue may be computed. The essential step in the algorithm is the evaluation of the unique positive solution of a quadratic matrix equation.

18 AFOSR

19 TR-78-1357

9 Interim rept.,

6

EFFICIENT ALGORITHMIC SOLUTIONS TO EXPONENTIAL TANDEM QUEUES WITH BLOCKING

by

10 Guy Latouche Marcel F. Neuts
Free University of Brussels and
University of Delaware

and

Marcel F. Neuts
University of Delaware

14 TR-78/12

16 2304

17 A5

Department of Statistics and Computer Science
Technical Report No. 78/12

11 Jun[redacted]978

12 31p.

This research was sponsored by the Air Force Office of Scientific Research Air Force Systems Command USAF, under Grant No. AFOSR-77-3236. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

15 AFOSR-77-3236

391 807

Abstract

Stable queueing systems consisting of two groups of servers, having exponential service times, placed in tandem and separated by a finite buffer, are shown to have a steady-state probability vector of matrix-geometric form. The queue is stable as long as the Poisson arrival rate does not exceed a critical value, which depends in a complicated manner on the service rates, the numbers of servers in each group, the size of the intermediate buffer and the unblocking rule followed when system becomes blocked. The critical input rate is determined in a unified manner.

For stable queues, it is shown how the stationary probability vector and other important features of the queue may be computed. The essential step in the algorithm is the evaluation of the unique positive solution of a quadratic matrix equation.

Key Words

Queueing systems, blocking, buffer models, computational probability

1. Introduction

The queueing model consisting of two units in series with a finite intermediate waiting room has an extensive literature, dating back to 1956 with the work of G. C. Hunt [9]. The study of blocking in two or more units in series without intermediate waiting spaces was initiated by B. Avi-Itzhak and M. Yadin [2]. Further contributions to this model are due to N. U. Prabhu [17] and A. B. Clarke [3].

Models in which there is a finite waiting room between the two units and the service times in the first unit have a general distribution were discussed by T. Suzuki [20], M. F. Neuts [11,12] and K. Hildebrand [7], basically using transform methods which are not readily computationally implemented. The thesis by I. Hannibalsson [5] utilizes a buffer model to represent a queue with delayed feedback. The second unit then represents a holding stage for those customers who will rejoin the queue in front of Unit I. In this paper and also in that by B. Wong, W. Giffin and R. L. Disney [22], the analysis of finite capacity buffer models is carried out by fairly involved spectral decompositions of the transition probability matrices. Related models, with finite total numbers of customers were treated in the papers by K. L. Arya [1] and O. P. Sharma [19]. These papers also do not have an algorithmic orientation.

In recent years, there has been a growing interest in the development of computational methods to evaluate the stationary probability vector and related quantities for tandem queues with blocking. This interest came primarily from the recognition that these models are useful in the study of the behavior of subsystems of computers. In addition to detailed descriptions of several computer-related applications, A. G. Konheim and M. Reiser [10] propose an algorithm for the solution of a system consisting of two single-server units with exponential service time distributions. They also allow feedback of some departures from the second server to the queue in front of the first unit. In [18], these same authors further considered more elaborate forms of feedback and discussed additional applications in computer modeling. Iterative numerical procedures of the Gauss-Seidel type, such as proposed by F. S. Hillier and R. W. Boling [8], may also be implemented for these models.

In addition, bounds on the blocking probability were investigated more recently by F. G. Foster and H. G. Perros [4]. A particularly detailed study of diffusion approximations in tandem queues is due to G. Newell [15,16]. The recent paper by J. M. Harrison [6] is also relevant in this context.

It is the purpose of this paper to show that a large number of buffer models with exponential servers may be numerically solved in a unified way. The key result identifies their stationary probability vector in a (modified) matrix-geometric form. Appropriately partitioning that vector $\underline{x}$ as $(\underline{x}_0, \underline{x}_1, \ldots)$, we show that

$$\underline{x}_i = \underline{x}_{r-1} R^{i-r+1}, \qquad \text{for } i \geq r-1,$$

where r is the number of servers in the first unit. The matrix R is the unique positive solution to a matrix quadratic equation. The spectral radius of R is less than one. The r vectors $\underline{x}_0, \ldots, \underline{x}_{r-1}$, are also uniquely determined.

The approach, which is used here, is already implicit in the thesis of V. Wallace [21], but the proofs are based on further refinements and generalizations given in Neuts [13,14].

Description of the Model

The system consists of units I and II and a finite intermediate buffer. Unit I consists of r parallel exponential servers, processing customers at the same rate α. In Unit II, c parallel exponential servers process customers at the common rate β. Arrivals to Unit I occur according to a homogeneous Poisson process of rate λ. (See Fig. 1.)

The servers in Unit II can be active as long as there are customers, who have completed a pass through Unit I and are requesting their service. There are $M-c-1 \geq 0$ places in the buffer, so that at most M-1 customers can be either waiting in the buffer or being processed by one of the servers in Unit II. If the number of customers who have completed a pass through Unit I but have not cleared Unit II reaches M, one of the servers in Unit I becomes blocked.

Depending on the application, the blocking of one or more servers in Unit I may affect the ability of the unblocked servers either to accept a customer for service or to complete a service in course. We shall assume that when the number of blocked servers in Unit I reaches r^*, $1 \leq r^* \leq r$, all unblocked servers in Unit I also cease service. This situation will be referred to as full blocking.

In a partially blocked system, when a service completion in Unit II occurs, one of the blocked servers of Unit I releases his customer into the buffer. This server may now again initiate a service.

Next we specify the unblocking rule. In a fully blocked system there are $M + r^* - 1$ customers who have completed a pass through Unit I and are requesting service in Unit II. We define an integer k^*, $0 \leq k^* \leq M+r^*-2$. When the number of customers, who have not been cleared by Unit II, drops to k^*, all interrupted services in the servers in Unit I resume and any free servers can again initiate services.

We shall allow a feedback loop of departures from Unit II back to the queue in front of Unit I. With probability $\theta' = 1 - \theta$, $0 \le \theta < 1$, a customer who completes a service in Unit II leaves the system. Feedback occurs with probability θ.

In order to concentrate only on parameters which have substantial significance, we shall not discuss further extensions in which customers may leave the system from Unit I or may enter feedback loops from the buffer to Unit I or from Unit II to the buffer. The relevant matrices which govern such cases can be constructed easily; the theorems and algorithms discussed below carry over routinely.

We also make the standard independence assumptions. All service times and interarrival times are mutually independent random variables. From a numerical viewpoint, it is routine to consider extensions such as the case where the rate of the Poisson arrival process depends on the number of blocked servers in Unit I, but in order not to add to the number of parameters of the model we shall not pursue this topic further.

Notational Convention

The material in this paper involves a large number of Jacobi matrices, whose detailed definitions require display. A matrix such as

$$\begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \end{array} \left| \begin{array}{cccccccc} b_0 & c_0 & 0 & 0 & & & & \\ a_1 & b_1 & c_1 & 0 & & & & \\ 0 & a_2 & b_2 & c_2 & & & & \\ & & & & \ddots & & & \\ & & & & a_{m-2} & b_{m-2} & c_{m-2} & 0 \\ & & & & 0 & a_{m-1} & b_{m-1} & c_{m-1} \\ & & & & 0 & 0 & a_m & b_m \end{array} \right|$$

will be displayed as

$$\left\| \begin{matrix} & c_0 & c_1 & \cdots & c_{m-3} & c_{m-2} & c_{m-1} \\ b_0 & b_1 & b_2 & \cdots & b_{m-2} & b_{m-1} & b_m \\ a_1 & a_2 & a_3 & \cdots & a_{m-1} & a_m & \end{matrix} \right\|$$

2. The Structure of the Markov Process

Under the assumption of exponential service times for the servers in Units I and II, the queueing model may be described as a continuous-parameter Markov chain on the state space $\{(i,j),\ i \geq 0,\ 0 \leq j \leq N\}$, where N is a finite nonnegative integer. The index i will denote the number of customers queued up or in service in Unit I. Such customers will be called I-customers. Upon completion of a pass through Unit I, a customer becomes a II-customer. We note that because of the possibility of feedback, a customer may be termed a I- or a II-customer several times in succession before leaving the system. The role played by the index j is more complicated to describe and will be spelled out for the specific cases discussed below.

In all cases, however, the infinitesimal generator P of the Markov chain will have the structure of a block-tridiagonal matrix of the form

$$(1) \quad P = \left\| \begin{matrix} & A_{02} & A_{12} & \cdots & A_{r-3,2} & A_{r-2,2} & A_2 & A_2 & \cdots \\ A_{01} & A_{11} & A_{21} & \cdots & A_{r-2,1} & A_{r-1,1} & A_1 & A_1 & \cdots \\ A_{10} & A_{20} & A_{30} & \cdots & A_{r-1,0} & A_0 & A_0 & A_0 & \cdots \end{matrix} \right\|$$

where all entries are square matrices of order N + 1. The rows in the block-partitioned matrix correspond to the sets of states $\{(i,0),(i,1),\ldots,(i,N)\}$, for $i \geq 0$.

We shall now give the detailed definitions of these blocks for various models of increasing complexity.

<u>Model A</u>

Unit I blocks as soon as there are M II-customers in the system. All c servers in Unit II are busy; there are M-c-1 customers in the waiting room and one server in Unit I has completed service of a customer who cannot enter the waiting room. Unblocking occurs as soon as a departure from Unit II occurs. In terms of our general description, Model A corresponds to $r^* = 1$, $k^* = M-1$.

In this case N = M. The matrices A_0, A_1 and A_2 are given by

$$A_0 = \left\| \begin{array}{cccccc} & r\alpha & r\alpha & \cdots & r\alpha & r\alpha \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \end{array} \right\|,$$

$$A_1 = \left\| \begin{array}{cccccccc} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ -\lambda-r\alpha & -\lambda-r\alpha-\beta & \cdots & -\lambda-r\alpha-(c-1)\beta & -\lambda-r\alpha-c\beta & \cdots & -\lambda-r\alpha-c\beta & -\lambda-c\beta \\ \beta\theta' & 2\beta\theta' & & c\beta\theta' & c\beta\theta' & & c\beta\theta' & \end{array} \right\|,$$

$$A_2 = \left\| \begin{array}{cccccccc} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \lambda & \lambda & \cdots & \lambda & \lambda & \cdots & \lambda & \lambda \\ \beta\theta & 2\beta\theta & \cdots & c\beta\theta & c\beta\theta & \cdots & c\beta\theta & \end{array} \right\|,$$

and for $1 \leq i \leq r-1$,

$$A_{i0} = \left\| \begin{array}{cccccc} & i\alpha & i\alpha & \cdots & i\alpha & i\alpha \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \end{array} \right\|.$$

For $0 \leq i \leq r-1$, the matrices $A_{i2} = A_2$, and the matrices A_{i1} are given by

$$A_{i1} = \left\| \begin{array}{cccccccc} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ * & * & \cdots & * & * & \cdots & * & * \\ \beta\theta' & 2\beta\theta' & \cdots & c\beta\theta' & c\beta\theta' & \cdots & c\beta\theta' & \end{array} \right\|.$$

The asterisks correspond to the negative diagonal entries, which are such that the row sums of the matrix P are zero.

Model B

This model is as the preceding one, except that full blocking occurs only when r^*, $1 \leq r^* \leq r$, servers in Unit I are blocked. The index j now ranges from 0 to $M + r^*-1$ and denotes the number of II-customers in the system. Unblocking occurs again upon a subsequent departure from Unit II, which corresponds to the case $k^* = M + r^*-2$.

The blocks in the partitioned matrix P are now of order $M + r^*$. They are obtained by augmenting the blocks in Model A in a systematic manner. Specifically

$$A_0 = \begin{array}{c} \left\| \begin{array}{ccccccc} \cdots & r\alpha & (r-1)\alpha & (r-2)\alpha & \cdots & (r-r^*+2)\alpha & (r-r^*+1)\alpha \\ \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & 0 & 0 & 0 & \cdots & 0 & \end{array} \right\| \\ \begin{array}{ccccccc} & \uparrow & & & & & \uparrow \\ & M & & & & & M+r^*-1 \end{array} \end{array}$$

$$A_1 = \begin{array}{c} \left\| \begin{array}{ccccccc} \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & * & * & * & \cdots & * & * \\ \cdots & c\beta\theta' & c\beta\theta' & c\beta\theta' & \cdots & c\beta\theta' & \end{array} \right\| \\ \begin{array}{ccccccc} & \uparrow & & & & & \uparrow \\ & M & & & & & M+r^*-1 \end{array} \end{array}$$

$$A_2 = \left\| \begin{array}{ccccccc} \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \lambda & \lambda & \lambda & \cdots & \lambda & \lambda \\ \cdots & c\beta\theta & c\beta\theta & c\beta\theta & \cdots & c\beta\theta & \end{array} \right\|$$

where the asterisks stand respectively for the entries $-\lambda-c\beta-(r-1)\alpha$, ..., $-\lambda-c\beta-(r-r^*+1)\alpha$, $-\lambda-c\beta$, chosen so that the row sums of P are zero.

The matrices A_{i0} are given by

$$A_{i0} = \begin{array}{c} \left\| \begin{array}{ccccccc} \cdots & i\alpha & \min(r-1,i)\alpha & \min(r-2,i)\alpha & \cdots & \min(r-r^*+1,i)\alpha \\ \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & 0 & 0 & 0 & \cdots & \end{array} \right\|; \\ \begin{array}{ccccccc} & \uparrow & & & & \uparrow \\ & M & & & & M+r^*-1 \end{array} \end{array}$$

for $1 \leq i \leq r-1$.

For $0 \leq i \leq r-1$, the matrices $A_{i2} = A_2$ and the matrices A_{i1} are given by

$$A_{i1} = \left\| \begin{array}{ccccccc} \cdots & 0 & 0 & \cdots & 0 & 0 \\ \cdots & * & * & \cdots & * & * \\ \cdots & c\beta\theta' & c\beta\theta' & & c\beta\theta' & \end{array} \right\|.$$

The asterisks correspond to the negative diagonal entries, which are such that the row sums in the matrix P are zero.

<u>Model C</u>

In this model, we add further complexity to Model B by assuming that when full blocking occurs, Unit I does not become unblocked until the number of II-customers drops to k^*. In most cases of interest, we will have $c \leq k^* \leq M+r^*-2$, and in order to limit the number of variants, we shall assume that this is the case. The case $k^* = M+r^*-2$ corresponds to Model B, so we only need to discuss the cases where $c \leq k^* \leq M+r^*-3$.

We now consider the indices j:

$0, 1, \ldots, M-1, M, \ldots, M+r^*-1, \overline{M+r^*-2}, \overline{M+r^*-3}, \ldots, \overline{k^*+1}.$

The index values with a bar correspond to the situations where the Unit I is blocked, although fewer than $M+r^*-1$ II-customers are in the system. The blocks in the partitioned matrix P are now matrices of order $2M+2r^*-k^*-2$.

The matrices A_0 and A_{i0}, $1 \leq i \leq r-1$, for this model are obtained by adding $M+r^*-k^*-2$ rows and columns to the corresponding matrices for Model B. These rows and columns are identically zero.

The matrices A_1 and A_{i1}, $0 \leq i \leq r-1$, are obtained by adding $M+r^*-k^*-2$ rows and columns to the corresponding matrices for Model B, and changing the row with index $M+r^*-1$. The diagonal elements to be added are all equal to $-\lambda-c\beta$. To the right of these diagonal elements we add an entry $c\beta\theta'$, except for the last row, where the entry $c\beta\theta'$ is placed in the column labeled k^*. In the row with index $M+r^*-1$ the entry $c\beta\theta'$ should appear immediately to the right, rather than to the left of the diagonal entry. All other added elements are zero.

The matrices A_2 and A_{i2}, $0 \leq i \leq r-1$, are similarly modified, with entries λ and $c\beta\theta$ now playing the role of the quantities $-\lambda-c\beta$ and $c\beta\theta'$.

We see that in the present model, the matrices are no longer Jacobi matrices, but remain highly structured sparse matrices. The theoretical results in this paper do not depend on the detailed structure of the blocks in the partitioned matrix P, but particularly when the order of these blocks becomes large, their sparsity may be exploited to economize on the storage and processing time requirements of the algorithm.

3. Quasi-Birth-and-Death Process

Consider an irreducible continuous parameter Markov chain with state space $\{(i,j);\ i \geq 0,\ 0 \leq j \leq N\}$ and infinitesimal generator P of the form (1).

Let us denote by $\underline{x}$ the vector of steady-state probabilities, associated to P, $\underline{x}\,P = \underline{0}$, $\underline{x}\,\underline{e} = 1$, and define the conservative stable matrix A by $A = A_0 + A_1 + A_2$. We assume that A is irreducible and denote by $\underline{\pi}$ its vector of steady-state probabilities, i.e. $\underline{\pi}\,A = \underline{0}$, $\underline{\pi}\,\underline{e} = 1$. Each component of $\underline{\pi}$ is strictly positive. In the tandem queue models considered here, A will obviously be irreducible.

We partition $\underline{x}$ as $\underline{x} = (\underline{x}_0, \underline{x}_1, \ldots)$, where each vector $\underline{x}_i$ has N+1 components. We shall examine below the existence of a solution of the form $\underline{x}_i = \underline{x}_{r-1} R^{i-r+1}$ for $i \geq r-1$, where R has a spectral radius strictly less than one (sp(R) < 1). For such a solution to exist, we must have that

$$\underline{x}_0 A_{0,1} + \underline{x}_1 A_{1,0} = \underline{0},$$

$$\underline{x}_i A_{i,2} + \underline{x}_{i+1} A_{i+1,1} + \underline{x}_{i+2} A_{i+2,0} = 0, \qquad \text{for } 0 \leq i \leq r-3,$$

$$\text{(2)} \quad \underline{x}_{r-2} A_{r-2,2} + \underline{x}_{r-1} A_{r-1,1} + \underline{x}_r A_0 = \underline{0},$$

$$\underline{x}_{r-1} R^{i-r+1} (A_2 + R A_1 + R^2 A_0) = \underline{0}, \qquad \text{for } i \geq r-1.$$

We shall show that in the positive recurrent case, a strictly positive probability vector $\underline{x}$ of the stated form exists, for which the matrix R is a nonnegative irreducible matrix of spectral radius sp(R) less than one and such that $A_2 + R A_1 + R^2 A_0 = 0$.

We now have to make several technical assumptions that are satisfied for the models we consider:

a. A_1 is nonsingular. By Wallace [21], Theorem 3.1, a sufficient condition is that $A_1 \underline{e} < \underline{0}$ which means that from any state (i,j), $i \geq r$, it is possible to move in one step to a state (i+1, j') or (i-1, j'). By Wallace [21], Lemma 3.4, A_1^{-1} is a nonpositive matrix with strictly negative diagonal elements.

b. The matrix $C_2 = -A_2 A_1^{-1}$, has at least one nonzero element in each row. A sufficient condition is that all diagonal entries of A_2 are strictly positive, which means that arrivals can occur when the system is in any state (i,j), $i \geq r$.

c. If we define $C_0 = -A_0 A_1^{-1}$ and $C = C_0 + C_2$, we assume that C is irreducible.

The equation $A_2 + R A_1 + R^2 A_0 = 0$, may now be rewritten in the form

$$\text{(3)} \quad R = C_2 + R^2 C_0.$$

Lemma 1

The matrix C has a maximal eigenvalue equal to one with corresponding left and right eigenvector, respectively proportional to $\underline{\pi}$ and $A_1 \underline{e}$.

Proof:

a. $\underline{\pi} C = -\underline{\pi} (A_0 + A_2) A_1^{-1} = \underline{\pi} (A_1 - A) A_1^{-1} = \underline{\pi}$,

since $\underline{\pi} A = \underline{0}$.

b. $C A_1 \underline{e} = -(A_0 + A_2) \underline{e} = (A_1 - A) \underline{e} = A_1 \underline{e}$,

since $A \underline{e} = \underline{0}$.

Define the sequence $\{R(n), n \geq 0\}$ of matrices as follows:

$R(0) = 0$,

$R(n+1) = C_2 + R(n)^2 C_0$, for $n \geq 0$.

Theorem 1

If $\underline{\pi} A_0 \underline{e} \leq \underline{\pi} A_2 \underline{e}$, the equation $R = C_2 + R^2 C_0$ has a unique solution R for which $R \geq 0$, $sp(R) \leq 1$. This solution is $\lim_{n \to \infty} R(n)$ and $sp(R) = 1$.

If $\underline{\pi} A_0 \underline{e} > \underline{\pi} A_2 \underline{e}$, the equation $R = C_2 + R^2 C_0$ has a unique solution R for which $R \geq 0$, $sp(R) < 1$. This solution, $\lim_{n \to \infty} R(n)$, is the minimal solution to the equation and is irreducible.

Remark

Before proving the theorem, we observe that assumptions b. and c. above are needed to prove that any matrix R satisfying equation (3) is irreducible. In each of the buffer models that we consider, one can show easily that in fact $\lim_{n \to \infty} R(n)$ is a strictly positive matrix.

Proof

This theorem is proved by repeating almost verbatim the argument given in [14], Theorems 1 and 2, Lemmas 2, 3 and 4. We only indicate the main steps of the proof here:

- For all n, $\underline{\pi} R(n) \leq \underline{\pi}$, hence $sp[R(n)] \leq 1$, for all n.

- For all n, $R(n+1) \geq R(n)$, therefore, the sequence converges monotonically to a matrix R such that

$$R = C_2 + R^2 C_0 \geq 0,$$

$$\underline{\pi} R \leq \underline{\pi}, \qquad sp(R) \leq 1.$$

- Any matrix $R \geq 0$, with $sp(R) \leq 1$, which satisfies $R = C_2 + R^2 C_0$, is irreducible; its maximum eigenvalue η is a solution of the equation $\eta = \theta(\eta)$, where for $0 < z \leq 1$, $\theta(z)$ is the maximal eigenvalue of the irreducible matrix $C(z) = C_2 + z^2 C_0$.

- The equation $z = \theta(z)$, $0 < z \leq 1$, has the unique solution $\eta = 1$, iff $\theta'(1-) \leq 1$. If $\theta'(1-) > 1$, that equation has two solutions: $\eta_1 = \eta < 1$ and $\eta_2 = 1$. Moreover, $\theta'(1-) = \underline{v}\, C'(1)\, \underline{u}$, where $\underline{v}$ and $\underline{u}$ are respectively the left and right Perron-Frobenius eigenvectors of C, normalized so that $\underline{v}\,\underline{e} = 1$, $\underline{v}\,\underline{u} = 1$. However, $\underline{v} = \underline{\pi}$ and $\underline{u} = (\underline{\pi} A_1 \underline{e})^{-1} A_1\underline{e}$, so that $\theta'(1-) = -2\, \underline{\pi} A_2 \underline{e}\, (\underline{\pi} A_1 \underline{e})^{-1}$. Clearly $\theta'(1-) \leq 1$, if and only if $-2\, \underline{\pi} A_2 \underline{e} \geq \underline{\pi} A_1 \underline{e}$, (as $A_1 \underline{e} \leq 0$) or equivalently,

$$\underline{\pi} A_2 \underline{e} \leq -\underline{\pi} (A_1+A_2) \underline{e} = -\underline{\pi} (A-A_0) \underline{e} = \underline{\pi} A_0 \underline{e}.$$

- The matrix $R = \lim_{n\to\infty} R(n)$ is the minimal solution to the equation, $R = C_2 + R^2 C_0$, in the set of nonnegative matrices with $sp(R) \leq 1$. This matrix has as its spectral radius the smallest positive root of $z = \theta(z)$, and is the only such matrix.

We now discuss the characterization of the vectors $\underline{x}_0, \ldots, \underline{x}_{r-1}$, in the case where $\underline{\pi} A_0 \underline{e} > \underline{\pi} A_2 \underline{e}$. By repeating an argument given in Neuts [13], one may show that this is the only case for which the Markov process P is positive recurrent.

Let $\underline{x}^* = (\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_{r-1})$, and

$$P^* = \left\| \begin{matrix} & A_{02} & \cdots & A_{r-3,2} & A_{r-2,2} \\ A_{01} & A_{11} & \cdots & A_{r-2,1} & A_{r-1,1} + R\,A_0 \\ A_{10} & A_{20} & \cdots & A_{r-1,0} & \end{matrix} \right\|$$

Lemma 2

P^* is an infinitesimal generator.

Proof

Since P is an infinitesimal generator and $R\,A_0 \geq 0$, all off-diagonal elements of P^* are non-negative.

To prove that $P^*\,\underline{e} = \underline{0}$, one needs only consider the last N+1 rows of P^*, since the other rows are identical to rows of P. However

$$A_{r-1,0}\,\underline{e} + (A_{r-1,1}+RA_0)\,\underline{e} = -A_2\,\underline{e} + R\,A_0\,\underline{e}$$

$$= -A_2\,\underline{e} + R\,A_0\,\underline{e} + \sum_{\nu=0}^{\infty} R^{\nu}\,(A_2+RA_1+R^2A_0)\,\underline{e}$$

$$= R(I-R)^{-1}\,(A_0+A_1+A_2)\,\underline{e} = \underline{0}.$$

Lemma 3

If $A_{r-1,1} + R\,A_0$ is irreducible, then P^* is irreducible.

Note that this condition is not necessary but obviously holds in the models we consider.

Proof

P^* is irreducible if and only if for all (i,j) and (i',j') (i,i' = 0,1, ..., r-1; j, j' = 0,1, ..., N) there exists in P^* a path from (i,j) to (i',j'). As P is irreducible, there exists in P a path for all such (i,j) and (i',j').

a. The path in P involves only states (i",j") with $0 \leq i" \leq r-1$. Then the same path exists in P^*.

b. The path in P involves some state (i'',j'') with $i'' \geq r$.

Let $(r-1,j_1)$ and $(r-1,j_2)$ be defined as follows: $(r-1,j_1)$ is the last state in the path before passage for the first time through a state (i^*,j^*) with $i^* \geq r$; $(r-1,j_2)$ is the first state in the path after passage for the last time through a state (i^+,j^+) with $i^+ \geq r$. By the structure of P, those states exist and in fact $i^* = i^+ = r$.

If $A_{r-1,1} + R\,A_0$ is irreducible, there exists a path from $(r-1,j_1)$ to $(r-1,j_2)$ involving only states $(r-1,j'')$ and, therefore, there exists in P^* a path from (i,j) to (i',j').

Theorem 2

Under the assumption of Lemma 3 and $\underline{\pi}\,A_0\,\underline{e} > \underline{\pi}\,A_2\,\underline{e}$, let $R \geq 0$ be the minimal solution of $R = C_2 + R^2\,C_0$. Let $\underline{x}^* = (\underline{x}_0^*, \underline{x}_1^*, \ldots, \underline{x}_{r-1}^*)$ be a solution of $\underline{x}^*\,P^* = \underline{0}$, then $\underline{x}^*$ has components all of the same sign. Furthermore $\underline{x}^*$ may be normalized by

$$(4) \qquad \sum_{\nu=0}^{r-2} \underline{x}_\nu^*\,\underline{e} + \underline{x}_{r-1}^*\,(I-R)^{-1}\,\underline{e} = 1.$$

The vector $\underline{x} = (\underline{x}_0, \underline{x}_1, \ldots)$ with

$$(5) \qquad \begin{aligned} \underline{x}_i &= \underline{x}_i^*, && \text{for } 0 \leq i \leq r-1, \\ \underline{x}_i &= \underline{x}_{r-1}^*\,R^{i-r+1}, && \text{for } r-1 \leq i, \end{aligned}$$

is the unique, strictly positive steady-state probability vector of the matrix P.

The proof is now obvious.

Remark

Since R is irreducible and $sp(R) < 1$, $(I-R)^{-1}$ exists and is strictly positive.

Corollary 1

(6) $R A_0 \underline{e} = A_2 \underline{e}$.

Proof

$$R = C_2 + R^2 C_0 = -A_2 A_1^{-1} - R^2 A_0 A_1^{-1},$$

hence

$$R A_1 \underline{e} = -A_2 \underline{e} - R^2 A_0 \underline{e},$$

and

$$-R(A_0+A_2) \underline{e} = -A_2 \underline{e} - R^2 A_0 \underline{e},$$

$$(I-R) A_2 \underline{e} = (I-R) R A_0 \underline{e}.$$

Since I-R is nonsingular, Formula (6) follows.

Remark

In the tandem queue models, considered here, $A_0 \underline{e}$ is the vector of rates of departure from Unit I when all servers are busy. $A_2 \underline{e}$ is the vector of rates of arrival to the queue in front of Unit I. This corollary shows that R plays a role similar to a traffic coefficient. In numerical computations, this relation serves usefully as an accuracy check on the evaluation of R.

4. Explicit Forms of the Equilibrium Condition

For the specific versions A, B and C of the buffer model, the equilibrium condition $\underline{\pi} A_0 \underline{e} > \underline{\pi} A_2 \underline{e}$, may be explicitly written in terms of the parameters of the model. Although the analytic expressions of these explicit forms are complicated, they all are of the general form

(7) $\lambda < (1-\theta)\, r\, \alpha\, \Psi$

where $0 < \Psi < 1$, and Ψ is a function of all the parameters of the model, except for λ and θ.

The quantity $(1-\theta)\, r\alpha$ is the critical input rate of a system consisting only of Unit I with a feedback probability θ. The entire right hand side of (7) may be interpreted as the critical input rate λ^* to the system under consideration. The dependence of λ^* on the various parameters of the model provides us with a readily accessible means of comparing the effects of buffer size and unblocking rules. It must be borne in mind, however, that queues for which λ is close to or equal to λ^* will exhibit the typical, frequently undesirable, long-range fluctuations inherent in near-critical queues.

Theorem 3

The vector $\underline{\pi}$ and the equilibrium condition $\underline{\pi}\, A_0\, \underline{e} > \underline{\pi}\, A_2\, \underline{e}$, are given by:

For Model A

$$(8)\quad \pi_0 = \left[\sum_{j=0}^{c-1} \frac{1}{j!}\left(\frac{r\alpha}{\beta}\right)^j + \frac{c^c}{c!}\sum_{j=c}^{M}\left(\frac{r\alpha}{c\beta}\right)^j\right]^{-1},$$

$$\pi_j = \frac{1}{j!}\left(\frac{r\alpha}{\beta}\right)^j \pi_0, \qquad \text{for } 1 \le j \le c,$$

$$\pi_j = \frac{c^c}{c!}\left(\frac{r\alpha}{c\beta}\right)^j \pi_0, \qquad \text{for } c \le j \le M.$$

$$(9)\quad \lambda < (1-\theta)\, r\alpha \sum_{j=0}^{M-1} \pi_j = (1-\theta)\, r\alpha\, (1-\pi_M).$$

For Model B $(2 \le r^* \le r)$

$$(10)\quad \pi_0 = \left[\sum_{j=0}^{c-1} \frac{1}{j!}\left(\frac{r\alpha}{\beta}\right)^j + \frac{c^c}{c!}\sum_{j=c}^{M}\left(\frac{r\alpha}{c\beta}\right)^j + \frac{c^c}{c!}\sum_{j=M+1}^{M+r^*-1}\left(\frac{r\alpha}{c\beta}\right)^j \prod_{\nu=1}^{j-M}\left(1-\frac{\nu}{r}\right)\right]^{-1},$$

$$\pi_j = \frac{1}{j!}\left(\frac{r\alpha}{\beta}\right)^j \pi_0, \qquad \text{for } 1 \le j \le c,$$

$$\pi_j = \frac{c^c}{c!}\left(\frac{r\alpha}{c\beta}\right)^j \pi_0, \qquad \text{for } c \le j \le M,$$

$$\pi_j = \frac{c^c}{c!}\left(\frac{r\alpha}{c\beta}\right)^j \prod_{\nu=1}^{j-M}\left(1-\frac{\nu}{r}\right)\pi_0, \qquad \text{for } M+1 \le j \le M+r^*-1.$$

$$(11)\quad \lambda < (1-\theta)\, r\alpha \left[\sum_{j=0}^{M-1} \pi_j + \sum_{j=1}^{r^*-1} \left(1 - \frac{j}{r}\right) \pi_{M+j-1}\right]$$

$$= (1-\theta)\, r\alpha \left(1 - \pi_{M+r^*-1} - \sum_{j=1}^{r^*-1} \frac{j}{r}\, \pi_{M+j-1}\right).$$

For Model C ($1 \le r^* \le r$, $c \le k^* \le M+r^*-2$)

We shall give detailed formulas only for the most useful case where $c \le k^* \le M-1$. The equations for the other cases are entirely similar. We denote by $\tilde{\pi}_{k^*+1}, \ldots, \tilde{\pi}_{M+r^*-2}$, the components of $\underline{\pi}$ corresponding to the indices $j = \overline{k^*+1}, \ldots, \overline{M+r^*-2}$. The explicit formulas for the components of $\underline{\pi}$ are uninspiringly complicated, but their numerical values may readily be computed by solving the linear equations

$$(12)\quad \begin{aligned}
\pi_j &= \frac{r\alpha}{j\beta}\, \pi_{j-1}, && \text{for } 1 \le j \le c,\\
\pi_j &= \frac{r\alpha}{c\beta}\, \pi_{j-1}, && \text{for } c \le j \le k^*,\\
\pi_j &= \frac{r\alpha}{c\beta}\, \pi_{j-1} - \tilde{\pi}_{k^*+1}, && \text{for } k^*+1 \le j \le M,\\
\pi_{M+j} &= \left(1 - \frac{j}{r}\right)\left(\frac{r\alpha}{c\beta}\right) \pi_{M+j-1} - \tilde{\pi}_{k^*+1}, && \text{for } 1 \le j \le r^*-2,\\
\pi_{M+r^*-1} &= \tilde{\pi}_{M+r^*-2} = \cdots = \tilde{\pi}_{k^*+1} = \frac{c^c}{c!}\left(\frac{r\alpha}{c\beta}\right)^{M+r^*-1} \prod_{\nu=1}^{r^*-1}\left(1 - \frac{\nu}{r}\right) \Phi^{-1}\, \pi_0, &&
\end{aligned}$$

where

$$\Phi = \sum_{h=1}^{r^*-2} \left(\frac{r\alpha}{c\beta}\right)^h \prod_{\nu=1}^{h} \left(1 - \frac{r^*-\nu}{r}\right) + \prod_{\nu=1}^{r^*-1}\left(1 - \frac{\nu}{r}\right) \sum_{h=0}^{M-k^*-1} \left(\frac{r\alpha}{c\beta}\right)^{h+r^*-1}.$$

Finally π_0 is obtained from the normalizing condition $\underline{\pi}\,\underline{e} = 1$.

$$(13)\quad \lambda < (1-\theta)\, r\alpha \left[\sum_{j=0}^{M-1} \pi_j + \sum_{j=1}^{r^*-1} \left(1 - \frac{j}{r}\right) \pi_{M+j-1}\right].$$

Proof

We shall only sketch the proof for Model C. The equations $\underline{\pi}\, A = \underline{0}$ may be written as

(14) $-r\alpha\,\pi_0 + \beta\pi_1 = 0,$

$$r\alpha\,\pi_{j-1} - (r\alpha + j\beta)\,\pi_j + (j+1)\beta\pi_{j+1} = 0, \qquad \text{for } 1 \le j \le c-1,$$

$$r\alpha\,\pi_{j-1} - (r\alpha + c\beta)\,\pi_j + c\beta\,\pi_{j+1} + \delta_{j,k^*}\,c\beta\,\tilde{\pi}_{k^*+1} = 0,$$
$$\text{for } c \le j \le M-1,$$

$$(r-j)\alpha\ \ \pi_{M+j-1} - \Big[(r-j-1)\alpha + c\beta\Big]\pi_{M+j} + c\beta\,\pi_{M+j+1} = 0,$$
$$\text{for } 0 \le j \le r^*-3,$$

$$(r-r^*+2)\alpha\,\pi_{M+r^*-3} - \Big[(r-r^*+1)\alpha + c\beta\Big]\pi_{M+r^*-2} = 0, \quad (\text{for } r^* \ge 2)$$

$$(r-r^*+1)\alpha\,\pi_{M+r^*-2} - c\beta\,\pi_{M+r^*-1} = 0,$$

$$\pi_{M+r^*-1} = \tilde{\pi}_{M+r^*-2} = \ldots = \tilde{\pi}_{k^*+1}.$$

These are clearly equivalent to

(15) $r\alpha\,\pi_{j-1} = \min(j,c)\beta\,\pi_j, \qquad \text{for } 1 \le j \le k^*,$

$$r\alpha\,\pi_{j-1} = c\beta\,\pi_j + c\beta\,\tilde{\pi}_{k^*+1}, \qquad \text{for } k^*+1 \le j \le M,$$

$$(r-j)\alpha\,\pi_{M+j-1} = c\beta\,\pi_{M+j} + c\beta\,\tilde{\pi}_{k^*+1}, \qquad \text{for } 1 \le j \le r^*-2,$$

$$(r-r^*+1)\alpha\,\pi_{M+r^*-2} = c\beta\,\pi_{M+r^*-1},$$

$$\pi_{M+r^*-1} = \tilde{\pi}_{M+r^*-2} = \ldots = \tilde{\pi}_{k^*+1}.$$

Equating the expression recursively computed for π_{M+r^*-1} with $\tilde{\pi}_{k^*+1}$, we obtain the stated formula relating $\tilde{\pi}_{k^*+1}$ and π_0.

The inequality $\underline{\pi}\,A_0\,\underline{e} > \underline{\pi}\,A_2\,\underline{e}$, is equivalent to

$$\lambda < \sum_{j=1}^{c} (r\alpha\pi_{j-1} - j\beta\theta\pi_j) + \sum_{j=c+1}^{M} (r\alpha\pi_{j-1} - c\beta\theta\pi_j)$$
$$+ \sum_{j=1}^{r^*-1} \Big[(r-j)\,\pi_{M+j-1} - c\beta\,\theta\,\pi_{M+j}\Big] - c\beta\,\theta\,(M+r^*-k^*-2),$$

and by using the equations (14), we obtain Formula (13).

Remarks

1. It is preferable not to write the geometric sums in (8) and (10) in closed forms, so that we do not have to write separate expressions for the case where $r\alpha = c\beta$.

2. For $r = c = 1$, and λ chosen, without loss of generality, to be equal to one, we obtain for Model A that the queue will be stable if and only if

$$\sum_{\nu=0}^{M} \left(\frac{\alpha}{\beta}\right)^{\nu} < (1-\theta)\alpha \sum_{\nu=0}^{M-1} \left(\frac{\alpha}{\beta}\right)^{\nu}.$$

This agrees, after elementary manipulations, with the conditions (2) for $\alpha \neq \beta$, and (3) for $\alpha = \beta$, stated in Thm 2 of Konheim and Reiser ([10], p. 334). A minor correction is, however, needed in the statement of that theorem. Condition (1), i.e. $(1-\theta) \min(\alpha,\beta) > 1$, is claimed to be the equilibrium condition for the system where $M = \infty$. As it is implied by one of the other conditions, depending on whether $\alpha \neq \beta$ or $\alpha = \beta$, its inclusion in the stability condition for finite M is clearly inappropriate.

5. The System Considered at Service Completions in Unit I

Upon considering the numbers of I- and II-customers immediately after service completions in Unit I, we obtain a Markov chain with the states (i,j), where $i \geq 0$ and $j = 1, \ldots, M+r^*-1$. In the interest of notational simplicity, we shall preserve the earlier state space, but note that since service completions during full blocking in Unit I are impossible, the states with $j = 0$ and the additional states corresponding to full blocking are ephemeral. Our formulas will correctly assign "steady-state probabilities" equal to zero to all such states and it will not be necessary to adjust the dimensions of the matrices which are involved.

Theorem 4

The stationary probability vector $\underline{z} = (\underline{z}_0, \underline{z}_1, \dots)$ of the embedded Markov chain at service completions in Unit I is given by

$$(16) \qquad \underline{z}_k = \tau\, \underline{x}_{k+1} A_{k+1,0}, \qquad \text{for } 0 \le k \le r-1,$$

$$\underline{z}_k = \tau\, \underline{x}_{k+1} A_0 = \tau\, \underline{x}_{r-1} R^{k-r+2} A_0, \qquad \text{for } k \ge r-1,$$

where τ is given by

$$(17) \qquad \tau = \left[\sum_{i=1}^{r-1} \underline{x}_i A_{i,0}\, \underline{e} + \underline{x}_{r-1} R (I-R)^{-1} A_0\, \underline{e} \right]^{-1}.$$

The zero components in the vectors $\underline{z}_k$ are ignored.

Proof

The formulas (16) are readily obtained by a conditioning argument for the elementary probabilities. The quantity $\tau^{-1}dt$ is clearly the elementary probability that a service completion occurs in $(t,t+dt)$ and the components of $\underline{x}_{k+1} A_{k+1,0}\, dt$ or $\underline{x}_{k+1} A_0\, dt$ are elementary probabilities of transitions of the type $(k+1,j) \to (k,j')$.

Corollary 2

The components with $j = M + \nu$, $0 \le \nu \le r^*-1$, of the vector

$$(18) \qquad \underline{Z} = \sum_{k=0}^{\infty} \underline{z}_k = \tau \left[\sum_{i=1}^{r-1} \underline{x}_i A_{i0} + \underline{x}_{r-1} R(I-R)^{-1} A_0 \right],$$

yield the stationary probabilities that upon completion of a service in Unit I, $\nu + 1$ servers are blocked in Unit I.

It would satisfy higher standards of rigor to set up explicitly the transition probability matrix of the embedded Markov chain and to verify that $\underline{z}$ is indeed its invariant vector. In order to avoid introducing a large amount of extra notation, we shall only do this for the case $r = 1$. In the process we shall also obtain a different formula for $\underline{Z}$, which is also easily implemented and therefore provides us with an accuracy check in numerical computations.

Completely elementary probability arguments yield that for $r = 1$, the transition probability matrix of the embedded chain is given by

$$(19)\qquad \bar{P} = \begin{vmatrix} \bar{B}_0 & \bar{B}_1 & \bar{B}_2 & \bar{B}_3 & \bar{B}_4 & \cdots \\ \bar{A}_0 & \bar{A}_1 & \bar{A}_2 & \bar{A}_3 & \bar{A}_4 & \cdots \\ 0 & \bar{A}_0 & \bar{A}_1 & \bar{A}_2 & \bar{A}_3 & \cdots \\ 0 & 0 & \bar{A}_0 & \bar{A}_1 & \bar{A}_2 & \cdots \\ 0 & 0 & 0 & \bar{A}_0 & \bar{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{vmatrix},$$

where $\bar{A}_n = (-A_1^{-1} A_2)^n (-A_1^{-1} A_0)$, and $\bar{B}_n = (-A_{01}^{-1} A_2) \bar{A}_n$, for $n \geq 0$.

The steady-state equations $\underline{z}\, \bar{P} = \underline{z}$, will be satisfied by the vectors defined in Formula (16) provided that

$$(20)\qquad \underline{x}_0 R^{k+1} A_0 = \underline{x}_0 R A_0 (-A_{01}^{-1} A_2) \bar{A}_k + \sum_{\nu=1}^{k+1} \underline{x}_0 R^{\nu+1} A_0 \bar{A}_{k+1-\nu} A_0, \quad \text{for } k \geq 0.$$

Since $\underline{x}_0 = \underline{x}_0 R (-A_0 A_{01}^{-1})$, and using the explicit form of the matrices $\bar{A}_n$, this equation may be equivalently rewritten as

$$(21)\qquad \underline{x}_0 \left[R^{k+1} - (-A_2 A_1^{-1})^{k+1} - \sum_{\nu=1}^{k+1} R^{\nu+1} (-A_0 A_1^{-1}) (-A_2 A_1^{-1})^{k+1-\nu} \right] A_0 = \underline{0}, \quad \text{for } k \geq 0.$$

In order to see that the matrix in square brackets in zero, write the equation $R = (-A_2 A_1^{-1}) + R^2(-A_0 A_1^{-1})$, $k+1$ times. Multiply the ν-th equation on the left by $R^{\nu-1}$ and on the right by $(-A_2 A_1^{-1})^{k-\nu+1}$ and sum up.

Finally the expression for τ is obtained from the normalizing equation $\underline{z}\, \underline{e} = 1$.

Corollary 3

In the case $r = 1$, the vector $\underline{Z}$ is also given by

$$\underline{Z} = (\underline{\pi}\, A_0\, \underline{e})^{-1}\, \underline{\pi}\, A_0 \tag{22}$$

$$- \tau\, \underline{x}_0\, (A_{01}+A_2)\, (A_1+A_2)^{-1}\, A_0 \left[A - (\underline{\pi}\, A_0\, \underline{e})^{-1}\, A_0\, \Pi\, A_0\right]^{-1} (A_1+A_2),$$

where Π is a matrix with $M + 1$ identical rows given by $\underline{\pi}$.

Proof

Adding the steady-state equations for the matrix $\overline{P}$, we obtain

$$\underline{Z}\left[I + (A_1+A_2)^{-1}\, A_0\right] = \underline{z}_0\, (-A_{01}^{-1}\, A_2) \left[-(A_1+A_2)^{-1}\, A_0\right] - \underline{z}_0 \left[-(A_1+A_2)^{-1} A_0\right]. \tag{23}$$

The matrix $-(A_1+A_2)^{-1}\, A_0$ is a stochastic matrix, whose first column is zero. All other elements are strictly positive. It has the left invariant vector $\underline{\pi}\, A_0$, whose first component is zero and all other components strictly positive. It now follows readily from the theory of finite Markov chains that the matrix

$$I + (A_1+A_2)^{-1}\, A_0 + (\underline{\pi} A_0 \underline{e})^{-1} \Pi\, A_0 = (A_1+A_2)^{-1} \left[A - (\underline{\pi} A_0 \underline{e})^{-1}\, A_0\, \Pi\, A_0\right], \tag{24}$$

is nonsingular. Adding $(\underline{\pi}\, A_0\, \underline{e})^{-1}\, \underline{Z}\, \Pi\, A_0 = (\underline{\pi}\, A_0\, \underline{e})^{-1}\, \underline{\pi}\, A_0$, to both sides of Equation (23) and replacing $\underline{x}_0\, R\, (-A_0\, A_{01}^{-1})$ by $\underline{x}_0$, we obtain the stated formula after routine matrix manipulations.

We note that the formula assigns the correct value zero to the first component of $\underline{Z}$. Verifying that the result so obtained agrees with $\underline{Z} = \tau\, \underline{x}_0\, R\, (I-R)^{-1}\, A_0$, (Cor. 2) provides a partial check on numerical computations.

6. Remarks on Numerical Computations

The solution, presented here, lends itself to a ready numerical implementation. Efficient programming, which takes the high degree of sparsity of the transition probability matrix into account, results in substantial savings in memory storage and execution times. This is particularly worthwhile when the program is to be used to study the design and control aspects of the model. In such studies, one or more parameters of the model need to vary over a range of values, which may require a substantial number of executions of the program. In view of the complicated dependence of the model on each of its parameters, detailed numerical studies appear to be the only way of obtaining the hard qualitative information needed in problems of design and optimization. Such a numerical implementation is currently being done. The qualitative results derived from it will be discussed elsewhere.

The first step, after ascertaining that the queue is stable, is to compute the matrix R. This may be done by successive substitutions in the equation $R = -A_2 A_1^{-1} - R^2 A_0 A_1^{-1}$, starting with $R = 0$. The relation $R A_0 \underline{e} = A_2 \underline{e}$, proved in Lemma 2, serves as an accuracy check.

If $r > 1$, the vectors $\underline{x}_0, \ldots, \underline{x}_{r-1}$, are computed by solving the system of linear equations, discussed in Theorem 2. Since the number of equations in that system may be very large, it is again desirable to take the special structure of its coefficient matrix into account. This may be done as follows. In the system

(25) $$\underline{x}_0 A_{01} + \underline{x}_1 A_{10} = \underline{0}$$

$$\underline{x}_{\nu-1} A_{\nu-1,2} + \underline{x}_\nu A_{\nu,1} + \underline{x}_{\nu+1} A_{\nu+1,0} = \underline{0}, \qquad \text{for } 1 \leq \nu \leq r-2,$$

$$\underline{x}_{r-2} A_{r-2,2} + \underline{x}_{r-1} (A_{r-1,1} + RA_0) = \underline{0},$$

the matrices $A_{\nu,2}$, $0 \leq \nu \leq r-2$, are clearly nonsingular, so that, using all but the first equation, we can write the vectors $\underline{x}_0, \ldots, \underline{x}_{r-2}$, as $\underline{x}_\nu = \underline{x}_{r-1} C_\nu^*$, $0 \leq \nu \leq r-2$, where the matrices C_ν^* are readily computed. The first equation now yields

(26) $$\underline{x}_{r-1} (C_0^* A_{01} + C_1^* A_{10}) = \underline{0},$$

which together with the normalizing condition

$$(27) \qquad \underline{x}_{r-1} \left[\sum_{\nu=0}^{r-2} C_\nu^* \underline{e} + (I-R)^{-1} \underline{e} \right] = 1,$$

uniquely determines the vector $\underline{x}_{r-1}$ and hence also the vectors $\underline{x}_0, \ldots, \underline{x}_{r-2}$.

If there is need to economize on memory storage, as when r and the order of the matrices are large, we can avoid storing the matrices C_ν, as they may be evaluated recursively. This can be done using only three arrays of size N x N and one linear array of length N. In the latter the vector $\sum_{\nu=0}^{r-2} C_\nu^* \underline{e}$ is accumulated. This does not significantly increase the processing time, as the systems of equations $\underline{x}_\nu A_{\nu 2} = \underline{d}$, where $\underline{d}$ is a known vector, are particularly easy to solve.

The simplifications, discussed above, are particularly striking when $\theta = 0$, (no feedback) as the matrices $A_{\nu 2}$, $0 \leq \nu \leq r-2$, are then scalar matrices.

The remaining computations of the vector $\underline{x}$, of various moments and of the marginal queue length densities, as well as the blocking probabilities are now entirely routine.

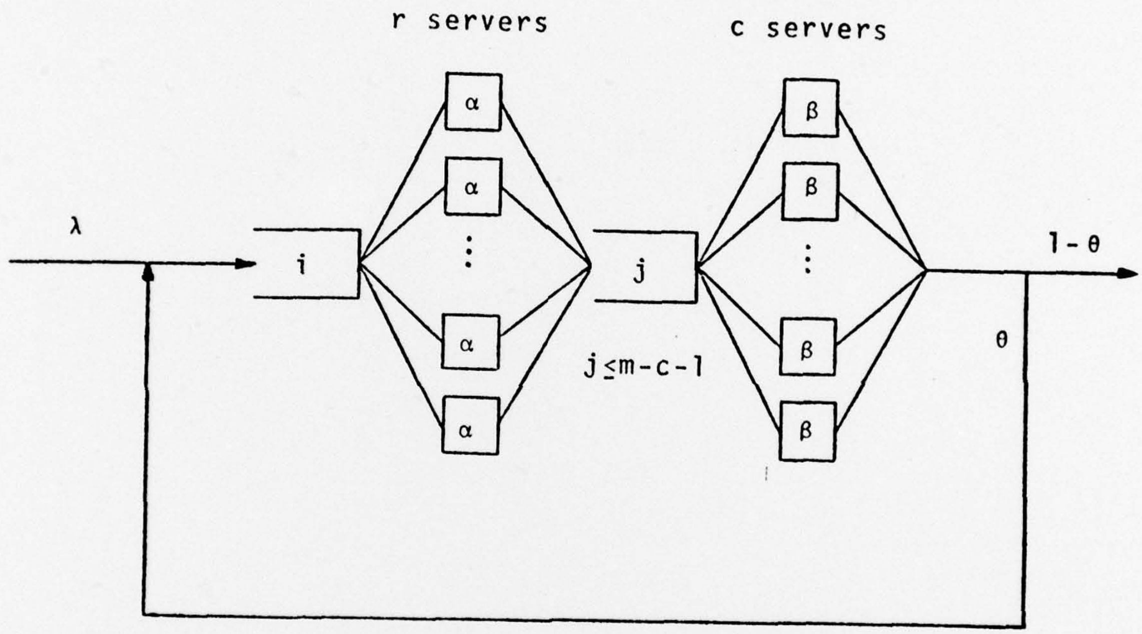

Figure 1

REFERENCES

[1] Arya, K. L. (1972)
Study of a Network of Serial and Non-Serial Servers with Phase Type Service and Finite Queueing Space
Journ. Appl. Prob., 9, 198-201.

[2] Avi-Itzhak, B. and Yadin, M. (1965)
A Sequence of Two Servers with no Intermediate Queue
Management Sci., 11, 553-564.

[3] Clarke, A. B. (1978)
Tandem Queues with Blocking
Abstract No. 160-35, Bull. Inst. Math. Stat., 7, 34.

[4] Foster, F. G. and Perros, H. G.
On the Blocking Process in Queue Networks
Manuscript, Dept. of Stat. Trinity College, Dublin, Ireland

[5] Hannibalsson, I. (1975)
Networks of Queues with Delayed Feedback
Technical Report No. 75-10, Dept. of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI

[6] Harrison, J. M. (1977)
The Diffusion Approximation for Tandem Queues in Heavy Traffic
Tech. Rept. No. 45, Dept. of Oper. Res., Stanford University, Stanford, CA.

[7] Hildebrand, D. K. (1967)
Stability of Finite Queue, Tandem Server Systems
Journ. Appl. Prob., 4, 571-583.

[8] Hillier, F. S. and Boling, R. W. (1971)
Finite Queues in Series with Exponential or Erlang Service Times - A Numerical Approach
Opns. Res., 15, 286-303.

[9] Hunt, G. C. (1956)
Sequential Arrays of Waiting Lines
Opns. Res., 4, 674-683.

[10] Konheim, A. G. and Reiser, M.
A Queueing Model with Finite Waiting Room and Blocking
J.A.C.M., 23, 328-341.

[11] Neuts, M. F. (1968)
Two Queues in Series with a Finite Intermediate Waitingroom
J. Appl. Prob., 5, 123-142.

[12] Neuts, M. F. (1970)
Two Servers in Series, treated in terms of a Markov Renewal Branching Process
Adv. Appl. Prob., 2, 110-149.

[13] Neuts, M. F. (1977)
The M/M/1 Queue with Randomly Varying Arrival and Service Rates
Tech. Rept. No. 77-11, Dept. of Stat. and Comp. Sci.
University of Delaware

[14] Neuts, M. F. (1978)
Markov Chains with Applications in Queueing Theory, which have a Matrix-geometric Invariant Vector
Adv. Appl. Prob., 10, 185-212.

[15] Newell, G. F. (1975)
Approximate Behavior of Tandem Queues (Chapters I - IV)
Research Report, Institute of Transportation and Traffic Engineering,
University of California, Berkeley CA

[16] Newell, G. F. (1977)
Approximate Behavior of Tandem Queues (Chapters V - VIII)
Research Report, Institute of Transportation Studies,
University of California, Berkeley CA.

[17] Prabhu, N. U. (1966)
Transient Behavior of a Tandem Queue
Management Sci., 13, 631-639.

[18] Reiser, M. and Konheim, A. G. (1978)
Finite Capacity Queueing Systems with Applications in Computer Modeling
SIAM Journ. on Computing, 7, 210-229.

[19] Sharma, O.P. (1973)
A Model for Queues in Series
Journ. Appl. Prob., 10, 691-696.

[20] Suzuki, T. (1964)
On a Tandem Queue with Blocking
Journ. Oper. Res. Soc. Japan, 6, 137-157.

[21] Wallace, V. (1969)
The Solution of Quasi Birth and Death Processes arising from Multiple Access Computer Systems
Ph.D. Thesis, Systems Engineering Lab, University of Michigan,
Ann Arbor, MI, Tech. Rept. No. 07742-6-T.

[22] Wong, B. Giffin, W., and Disney, R. L. (1977)
Two Finite M/M/1 Queues in Tandem: A Matrix Solution for the Steady State
Opsearch, 14, 1-18.